PREPRINT

Final peer-revised version

Global monitoring of soil multifunctionality in drylands using satellite imagery and field data

Hernández-Clemente R[1,2]*., Hornero A[3,4], Gonzalez-Dugo V[3], Berdugo M[5,6], Quero JL[1], Jiménez JC[7], Maestre FT[8,9].

Affiliations:

[1] University of Cordoba, Department of Forestry, Campus de Rabanales, Ctra. Madrid, Km 396, 14071, Cordoba, Spain.

[2] Swansea University, Department of Geography, Singleton Park, Swansea, SA2 8PP, UK.

[3] Instituto de Agricultura Sostenible (IAS), Consejo Superior de Investigaciones Científicas (CSIC), Avenida Menéndez Pidal s/n, 14004 Cordoba, Spain

[4] School of Agriculture and Food (SAF-FVAS) and Faculty of Engineering and Information Technology (IE-FEIT), University of Melbourne, 700 Swanston St, Carlton VIC 3053, Melbourne, Victoria, Australia

[5] ETH Zurich, Crowther Lab, CHN G 66, Universitätstrasse 16. 8092, Zürich.

[6] Departamento de Biodiversidad, Ecología y Evolución, Universidad Complutense de Madrid, Av. Séneca, 2, 28040, Madrid, Spain.

[7] GCU/IPL, University of Valencia, Catedrático José Beltrán 2, 46980 Paterna, Valencia, Spain
[8] Instituto Multidisciplinar para el Estudio del Medio "Ramón Margalef", Universidad de Alicante, Carr. de San Vicente del Raspeig, s/n, 03690, Alicante, Spain.
[9] Departamento de Ecología, Universidad de Alicante, Carr. de San Vicente del Raspeig, s/n, 03690, Alicante, Spain.

*Corresponding author: rociohc@uco.es

***Abstract*—** Models derived from satellite image data are needed to monitor the status of terrestrial ecosystems across large spatial scales. However, a remote sensing-based approach to quantify soil multifunctionality at the global scale is missing despite significant research efforts on this topic. A major constraint for doing so is the availability of suitable global-scale field data to calibrate remote sensing indicators (RSI) and, to a lesser extent, the sensitivity of spectral data of available satellite sensors to soil background and atmospheric conditions. Here, we aimed to develop a soil multifunctionality model to monitor global drylands coupling ground data on 14 soil functions of 222 dryland areas from six continents to 18 RSI derived from a time series (2006-2013) Landsat dataset. Among the RSI evaluated, the chlorophyll absorption ratio index was the best predictor of soil multifunctionality in single-variable-based models ($r=0.66$, $p<0.01$, NMRSE=0.17). However, a multi-variable RSI model combining the chlorophyll absorption ratio index, the global environment monitoring index and the canopy-air temperature difference improved the accuracy of quantifying soil multifunctionality ($r=0.73$, $p<0.01$, NMRSE=0.15). Furthermore, the correlation between RSI and soil variables shows a wide range of accuracy with upper and lower values obtained for AMI ($r=0.889$, NMRSE=0.05) and BGL ($r=0.685$, NMRSE=0.18), respectively. Our results provide new insights on assessing soil multifunctionality using RSI that may help to monitor temporal changes in the functioning of global drylands effectively.

Index Terms— Soil multifunctionality, global monitoring, satellite data, drylands, artificial intelligence.

Introduction

Drylands, areas with a precipitation/potential evapotranspiration ratio below 0.65 (Huang et al., 2016), are essential for sustaining life on our planet, as they cover around 42% of the global land surface, produce 42% of the world's food and host 30% of the world's endangered species (Gaur and Squires, 2017). However, drylands are threatened by climate change and desertification (Burrell et al., 2020), which can induce abrupt changes in their structure and functioning. These changes have been associated with increased aridity conditions (Berdugo et al., 2020) or reduced soil fertility and multifunctionality (Berdugo et al., 2017). Soil multifunctionality is understood as the ability of soils to maintain several ecosystem functions and services simultaneously (Garland et al., 2021). Consequently, it is crucial to monitor attributes of ecosystems, such as soil multifunctionality, in order to anticipate sudden changes that may be brought about by land degradation and the effects of climate change

Earth observation satellites are critical for monitoring temporal trends in ecosystem attributes across global drylands. Optical sensors with coarse spatial resolution, such as the National Oceanic and Atmospheric Administration Advanced Very High-Resolution Radiometer or satellite passive microwave observation, have provided valuable information on quantifying dryland biomass at the regional scale (Tian et al., 2016). However, empirically validating the data from these sensors is challenging because it requires measuring similar areas to their pixel sizes (>10 km^2). Broad-scale high-temporal frequency satellite data such as Landsat or MODIS have played an essential role in monitoring dryland vegetation dynamics. They have extensive spatial coverage and frequent observations, making them useful for this purpose. Landsat has been particularly successful in monitoring dryland vegetation attributes, with reliable accuracy in retrieving fractional cover and leaf area index at the regional scale (Sonnenschein et al., 2011; Sun, 2015). One of the methods most widely used to infer vegetation attributes has been the calculation of remote sensing indicators (RSI), such as the normalised difference vegetation index (NDVI)

(Rouse et al., 1974). However, NDVI applicability on a global scale is limited due to the spectral influence of mixed sparse vegetation and bare soil (Huete and Jackson, 1987).

Remote sensing indicators that minimise soil background have recently received considerable attention. Indices such as the soil-adjusted vegetation index (SAVI) (Huete et al., 1985), the optimised soil-adjusted vegetation index (OSAVI) (Rondeaux et al., 1996), the atmospherically resistant vegetation index (ARVI) (Kaufman and Tanre, 1992), the modified chlorophyll absorption in reflectance index (MCARI) (Haboudane et al., 2004a) or the global environment monitoring index (GEMI) (Ren and Feng, 2015) are more resistant than NDVI to saturation, background reflectance conditions and atmospheric effects. For instance, ARVI has a similar dynamic range to NDVI, but on average, it is four times less sensitive to atmospheric effects than NDVI (Kaufman and Tanre, 1992). However, the sensitivity of vegetation indices has mainly been studied at local and regional scales, and no studies have evaluated their suitability across global drylands. For instance, some studies show that traditional indices like NDVI perform better than modified vegetation indices for monitoring above-ground green biomass in arid and semi-arid grasslands (Ren and Feng, 2015). In contrast, other studies show that the modified vegetation indices, such as the SAVI and L-SAVI, improved the detection of spatio-temporal changes in the vegetation in a semi-arid area (Fatiha et al., 2013). These examples underscore the current deficiency in assessing and comparing existing vegetation indices at the global scale.

Developing global models of soil multifunctionality faces significant challenges, one of which is the lack of suitable field data that fits the spatial resolution of satellite imagery required to build and validate these models. To overcome this challenge, we need to move beyond remote sensing indicators (RSI) mainly related to plant cover and incorporate other indicators that can potentially analyze biophysical properties such as plant composition and functioning. For instance, Zhao et al. (2018) demonstrated a significant relationship between visible black-sky albedo and soil multifunctionality across global drylands. However, this study was limited to a selection of 61 homogeneous plots from the 224 dryland datasets compiled by Maestre et al. (2012) to avoid the mismatch between field data collected from 30 m × 30 m plots and MODIS image resolution of 500 m × 500 m (NASA LP DAAC, 2017). Such limitations can be overcome by combining existing global datasets of ground-collected soil data collected within 30 x 30 m plots (Delgado-Baquerizo et al., 2013; Maestre et al., 2012) with spectral data provided by Landsat 7 ETM. Landsat satellites offer the highest spatial resolution in the thermal region, with freely available imagery and the longest temporal record (roughly every 16 days) spanning the last 49 years (Wulder et al., 2022).

In comparison, new missions such as Sentinel-3 cover this region with a 1,000-m spatial resolution. Landsat features spatial resolutions ranging from 15 to 80 m in the visible and infrared region and between 60 to 120 m in the thermal region, depending on the specific Landsat mission (Holden and Woodcock, 2016). Integrating medium spatial resolution image data (30 m × 30 m) such as Landsat data and field-based observations could improve the assessment of soil multifunctionality worldwide in a cost-effective and accessible manner. Furthermore, the high-performance computational capacities of Google Earth Engine can be used to access and process satellite data on the cloud, providing new possibilities for analyzing large volumes of data globally (Gorelick et al., 2017).

Here, we combine the use of 18 surface reflectance vegetation indices and thermal remote sensing-based indices, hereafter called remote sensing indicators (RSI), with a global assessment of 14 soil functions measured *in situ* in 222 dryland ecosystems on six continents (Maestre et al., 2012; Ochoa-Hueso et al., 2018). Our study has two main objectives: firstly, to evaluate the sensitivity of remote sensing indicators (RSI) in characterizing soil multifunctionality in dryland ecosystems worldwide, and secondly, to develop models to upscale ground-based observations with RSI data. By achieving these objectives, we aim to provide robust and comprehensive models that can enhance our understanding of dryland ecosystems' current status and dynamics globally.

Material and methods

A. Study sites

Field data were gathered from 222 sites in 19 countries (Argentina, Australia, Brazil, Chile, China, Ecuador, Iran, Israel, Kenya, Mexico, Morocco, Peru, Spain, Tunisia, USA, Venezuela, Botswana, Burkina Faso and Ghana; (Fig. 1). These sites are a subset of the 236 sites used in Ochoa-Hueso et al. (2018); we had to exclude 14 sites due to the lack of cloud-free images during the inventories at these sites. The 222 sites surveyed covered all major vegetation/soil types and the wide range of environmental conditions across global drylands (UNEP-WCMC, 2007) (Fig. S1-Supplementary Material).

At each site, field data were collected from 30 × 30 m plots between February 2006 and December 2013 using a standardised protocol described in detail in Maestre et al. (2012). Plant cover data were obtained from four 30 m-long transects using the line-intercept method (Tongway and Hindley, 2004). Soil samples were collected using a stratified procedure of five 50 x 50 cm quadrats randomly placed under the dominant perennial vegetation patch type and in open areas devoid of perennial vegetation. Five soil cores extracted at 0–7.5 cm depth from each quadrat were bulked and homogenised in the field. The composite included samples for microsites in open areas devoid of perennial vegetation and under the canopy of the dominant perennial plant species. The number of soil samples collected varied between 10 and 15 per site (depending on whether one or two dominant plant species were found at each site), accounting for more than 2600 samples. After field collection, the soil samples were taken to the laboratory, where they were sieved (2 mm mesh), air-dried for one month, and stored for laboratory analysis. Dry soil samples were then analysed for soil functions related to the cycling and storage of carbon (Table 1), as described in Maestre et al. (2012). A plot-level estimate of all the soil functions was obtained using a weighted average of values from open and vegetated microsites weighted by their respective cover at each plot. As a soil multifunctionality index, we used the average Z-score for all soil functions estimated at the plot level (Maestre et al., 2012).

B. Satellite data and processing

We used Landsat 5 TM and Landsat 7 ETM+ filtered products to obtain spectral data from 927 available images collected between 2006 and 2013. First, we selected images taken as close as possible (within 1-3 months) to the day when field surveys were conducted for the 236 sites analysed in Ochoa-Hueso et al. (2018). We then used a local filter to select the cloud-free Landsat images closest to the field surveys, reducing the dataset to 222 sites.

To correct for atmospheric gases and aerosols, which can vary in space and time and can significantly impact Landsat spectral data collected on different dates (Masek et al., 2006; Roy et al., 2014), we atmospherically corrected the Landsat imagery using the Landsat ecosystem disturbance adaptive processing system (LEDAPS, version 3.4.0) (Schmidt et al., 2013). We also corrected surface reflectance to account for data from plots measured on different dates. To retrieve surface temperature, we used a methodology proposed by Jimenez-Muñoz et al. (2009) that employs a single-channel algorithm using the thermal-infrared Landsat channel (band 6). This algorithm conducts emissivity and atmospheric corrections to retrieve the surface-level temperature. The algorithm has been extensively validated in other independent studies across various land covers, showing an RMSE between 1 and 2 K, which is typical accuracy for remotely sensed land surface temperature products (Copertino, 2012; Z. Zhang et al., 2016). First, we estimated surface emissivity using a simple approach based on fractional vegetation cover and NDVI (Sobrino et al., 2008). We then calculated atmospheric functions from total atmospheric water vapour values obtained from the Copernicus Climate Change Service implemented by the European Centre for Medium-Range Weather Forecasts (Muñoz-Sabater et al., 2021). The single-channel algorithm needed these data to be applied (Hersbach et al., 2020). We applied a 30-metre buffer to extract data from each study area and weighted the value of each pixel covered in this area to minimize errors in the geolocation and referencing of each pixel. An average of the pixel weights by the percentage of the area overlapping each plot was used to ensure an extensive, systematically collected sample scheme.

We obtained spectral reflectance for each TM and ETM+ reflective band and surface temperature, which we used to calculate the RSI dataset for each location. To estimate a wide range of soil and plant traits (Hernández-Clemente et al., 2019), we evaluated a list of 18 RSI, including formulations based on the near-infrared and visible regions (NDVI, GLI, SIPI), modified vegetation indices proposed to minimize background and soil effects (SAVI, TSAVI, OSAVI, TSAVI/OSAVI, MCARI2 and GEMI), modified vegetation indices considering atmospheric corrections (ARVI, AFRI and VARI), formulations based on the short-wave infrared (SWIR) bands (S1260 and NBR2), and thermal bands (Ts-Ta) and WDI. For definitions and descriptions of acronyms, refer to Table 2.

C. Modelling approach

In this study, we investigated the capacity of the RSI evaluated to predict soil multifunctionality across global drylands. The first step in data analysis entailed selecting the most significant RSI for determining soil multifunctionality, followed by an evaluation of the model's performance, as depicted in Fig. 2. To ensure the interpretability of our model output, we first reduced the number of variables by using a filter-based feature selection approach (Gosiewska et al., 2021). We excluded RSI that were highly correlated with each other ($r > 0.85$, (Dormann et al., 2013)) and used only those with a variance inflation factor lower than ten (Kutner et al., 2004). This resulted in an RSI selection referred to as RSI-mc. We then performed a principal component analysis (PCA) to interpret the contribution of each RSI based on the first two principal components with an importance higher than 15% (Pacheco et al., 2013). We identified the loading vectors in the biplot of the principal components explaining >60% of the variance, which were used to select the RSI variables with the highest eigenvalues per axis, resulting in an RSI selection referred to as RSI-pca. Lastly, we included a single-index selection using the RSI most correlated with soil multifunctionality, referred to as the 1-RSI model, to check the improvement achieved with the variable reduction approaches (RSI-mc and RSI-pca).

To evaluate the suitability of our model, we compared three different approaches: two artificial intelligence methods, an evolutionary algorithm model (EAM) and a random forest model (RF), and simple linear regression (LR). The comparison between these three methods was made to analyse the suitability of each approach, which varies with the variability and dispersion of the data (Franklin and Miller, 2010). The EAM model was based on a genetic algorithm used to generate high-quality solutions to optimise model accuracy in computer science, known as evolutionary algorithms (Vikhar, 2016). We computed the models with the Eureqa software package v1.24 (Datarobot Inc, Boston, USA), combining the arithmetic, trigonometric and exponential building blocks for the best model accuracy. Eureqa uses evolutionarily search to determine the best predictive models, simplifying the final calculated model. The RF model was built by using the Caret library (Kuhn et al., 2020) within the R environment (R Core Team, 2013) and the package "caret". The adjustment parameter *mtry* (Randomly Selected Predictors) was established initially by iterating over the whole range of values. Then, a pre-processing transformation was applied by centring and scaling the training data. Comparatively, we also tested a simple method based on a linear fitting (Freedman, 2009) between RSI and soil multifunctionality. Finally, the models were trained considering a resampling method of five k-folds and three repetitions. The model accuracy was evaluated with a cross-validation bootstrap procedure (Austin and Tu, 2004). For doing so, data were randomly split into K=500 sets, selecting 80% of our dataset to generate each predictive model, and the remaining 20% was set aside for validation purposes. The average and standard deviation from this cross-validation bootstrap procedure was used for validation. We calculated the R-squared (R^2) and the normalised root-mean-square error (NRMSE) by contrasting predicted versus observed values.

Results

A. Soil multifunctionality RSI determination

The feature reduction simplified the number of variables included in the modelling process. For example, the first reduction, RSI-mc, resulted in a list of ten RSI: MCARI2, NBR2, MSAVI, GLI, S1260, AFRI22, TSAVI_OSAVI, GEMI, Ts-Ta, and WDI. In the PCA biplot, we observed that these RSI were grouped into three clusters of loading vectors, each associated with climate (Fig. 3a) and vegetation type (Fig. 3b) and enclosed by concentration ellipses.

The modified vegetation indices MCARI2, MSAVI, GLI, TSAVI_OSAVI, S1260, NBR2, and AFRI22 were negatively correlated to the same principal component (PC1), while GEMI was positively related to PC1. On an orthogonal axis, the third group of vectors, Ts-ta and WDI, were the best contributors to PC2. In the PCA biplot, the contribution of WDI and Ts-Ta was quite similar, with the eigenvalue being slightly higher for Ts-Ta. However, it should be noted that WDI was significantly more correlated than Ts-Ta with the soil functions evaluated (Fig. 4).

The ellipses in the Standardized Principal Components (PC1 vs PC2) plot serve as visual representations of the distribution and variability of data points associated with the same climate (Fig. 3a) and vegetation type (Fig. 3b). In these plots (Fig. 3a and b), large ellipses centered within the graph represent semi-arid climates, grasslands, and open forests. These ellipses demonstrate a high degree of variability in the data for each cluster and a consistent representation across the ten RSI-mc selected. Conversely, certain ellipses are associated explicitly with particular RSI. For example, GEMI and water stress indicators (Ts-Ta and WDI) contribute more to sites in arid areas. At the same time, GEMI and soil-adjusted, atmospherically resistant RSI (MCARI2, MSAVI, GLI, TSAVI_OSAVI, S1260, NBR2, and AFRI22) are more prominent in dry-subhumid areas (Fig. 3a and S1 - Supplementary Material). In the standardized principal components plot of vegetation types (Fig. 3b), GEMI and water stress indicators mainly represent savannahs. Shrublands form a cluster to the left, characterized by soil-adjusted, atmospherically resistant RSI and water stress indicators (Fig. 3b).

We selected one RSI per group with the highest eigenvalues from the three main groups of eigenvectors in the standardized principal components (PC1 vs PC2) plot (Fig. 3). As a result, RSI-pca selection reduced the predictors to MCARI2, GEMI, and Ts-Ta. Finally, we compared the RSI reductions, RSI-mc and RSI-pca, to the 1-RSI with the highest correlation for estimating soil multifunctionality. According to Fig. 4, MCARI2 strongly correlates with soil multifunctionality ($R=0.54$, $R^2=0.28$, $p<0.01$).

B. Global models of soil multifunctionality

The scaling-up of soil multifunctionality across a wide range of climates and vegetation types with

EAM, RF and LM models showed R^2 and NMRSE values ranging from 0.17 to 0.55 and from 0.25 to 0.15, respectively, which depended on the type of model and number of RSI considered (Fig. 5a). The highest accuracy for soil multifunctionality was obtained with models computed with EAM using the RSI-pca selection, which improved NRMSE by 25 % from RF and 27% from LM models, respectively (Fig. 5b). The EAM analysis reduced the NMRSE in soil multifunctionality estimations for the three variable selection methods followed (RSI-pca, RSI-mc and 1-RSI). The EAM-driven analysis utilizing MCARI2 resulted in a 22% reduction in NMRSE compared to the linear analysis derived from RSI-pca. Furthermore, it was observed that employing models based on RSI-mc was unnecessary, as the RSI-pca produced the most reliable outcomes when processed through EAM (Fig. 5a).

The accuracy of the soil multifunctionality model based on RSI-pca using EAM analyses (r=0.733, NMRSE=0.15) also shows consistency across soil functions, with upper and lower values obtained for AMI (r=0.889, NMRSE=0.05) and BGL (r=0.685, NMRSE=0.18), respectively (Fig. 6).

Discussion

We developed and validated a model to estimate soil multifunctionality across global drylands using a comprehensive global field survey and satellite imagery. Our results highlight the reliability of RSI, such as MCARI2, NBR2, MSAVI, GLI, S1260, AFRI22, TSAVI_OSAVI, or GEMI, to model dryland soil multifunctionality. These RSI were developed to reduce the influence of soil background and atmospheric effects on the regions with low-density vegetation cover. In contrast, other simpler RSI formulations, such as the widely used NDVI, showed about 50% lower coefficient of determination values than MCARI2 (see Fig. S2.- Supplementary Material) to model soil multifunctionality. These results may be related to a higher sensitivity of NDVI to soil brightness effects or to the presence of senesced vegetation and standing litter (Baret et al., 1993). However, NDVI could still provide complementary information to modelling soil multifunctionality with MCARI2 as a proxy for primary production (Prince, 1991; Tucker et al., 1983) or ecosystem structure and functioning (Gaitán et al., 2013) over large scales. In contrast to studies mainly based on the relationship between fractional vegetation cover and NDVI (Song et al., 2017), here we developed a model to monitor soil multifunctionality, a key feature of dryland ecosystems (Maestre et al., 2016).

The models based on the combination of MCARI2, GEMI and Ts-Ta (RSI-pca) improved the accuracy in estimating soil multifunctionality compared to models using a single predictor. While Ts-Ta alone cannot serve as a reliable indicator of soil multifunctionality, it can complement other RSI, such as MCARI2 and GEMI, to enhance the accuracy of global models. When combined with MCARI2 and GEMI, Ts-Ta improves the accuracy of soil multifunctionality models, as demonstrated by a 12% decrease in NMRSE, and enhances the prediction of specific soil functions by 8-18% (Table S1). This outcome may be attributed to the combination of RSI-pca selections (MCARI2, GEMI, and Ts-Ta), which more accurately represent the variability observed in drylands distributed across diverse climates and vegetation types worldwide. In addition to the strong correlations found between MCARI2, GEMI, and Ts-Ta and soil multifunctionality, further analyses using structural equation modelling (Table S2 and Fig. S3) demonstrate that these RSI provide the most reliable models for estimating soil multifunctionality in drylands. These findings align with prior research, indicating that MCARI2 and GEMI indices (Haboudane et al., 2004b; Pinty and Verstraete, 1992) exhibit lower sensitivity compared to other vegetation indices such as

NDVI in detecting fractional cover variations ranging from 2% to 83% across analyzed dryland locations (Maestre et al., 2012).

The RSI-pca model improves estimates of soil multifunctionality and individual soil functions. While the correlations of the 1-RSI model are significantly related to most of the soil functions (TON, BGL, ORC, PRO, PHE, ARO, HEX, and NTR), the RSI-pca model can generate models with errors of only 5-18% for all variables. This can be explained by the indirect relationship that many soil components have with different types of variables, such as variations in biomass, soil moisture, and primary production (Liu et al., 2023). The combined use of a model that absorbs this variability can reflect the specific variations of these compounds, as demonstrated in this work for AVP, NIT, AMI, and PEN.

Our study emphasizes the importance of avoiding models based solely on best-fitting indices (Hornero et al., 2021). The PCA reduction method improves the results' interpretability by evaluating the RSI loading vectors used to assess soil multifunctionality and functions per climate and vegetation type. Among the selected RSI, MCARI2 and GEMI are used as a proxy for fractional cover (Haboudane et al., 2002; Pinty and Verstraete, 1992), where MCARI2 reduces the RSI's sensitivity to soil and background effects, and GEMI minimizes the impact of undesirable atmospheric perturbations. Additionally, Ts-Ta provides an indicator of the water stress condition of the vegetation linked to stomatal conductance and transpiration (Morillas et al., 2013), contributing to representing semi-arid dryland sites.

Our study provides compelling evidence that EAM methods are a reliable tool for accurately upscaling ground-based observations of soil multifunctionality on a global scale. The EAM models developed in this study showed significant improvement in NRMSE values by 37% and 33%, respectively, compared to RF and LR models for quantifying soil multifunctionality

(Table S1). Furthermore, the accuracy obtained for predicting soil multifunctionality using the 1-RSI ($r=0.66$, $p<0.01$) and RSI-pca ($r=0.73$, $p<0.01$) models with Landsat data and EAM models represents a significant improvement compared to results from previous studies. For instance, Zhao et al. (2018) reported a correlation between soil multifunctionality and MODIS land surface albedo of only $r= -0.314$. These findings align with recent efforts to apply deep learning approaches to quantify soil organic carbon composition at the national level, as reported by Odebiri et al. (2022). These results demonstrate the potential of EAM models for providing reliable estimates of soil multifunctionality and support their application for global-scale monitoring and management of soil resources in drylands.

Biocrusts are essential components of drylands globally, significantly regulating their structure and functioning (Bowker et al., 2013; Maestre et al., 2013, 2011). Biocrusts fix substantial amounts of atmospheric CO_2 (over 2.6 Pg of C per year) (Elbert et al., 2012) and impact the temporal dynamics of soil CO_2 efflux and net CO_2 uptake. Additionally, biocrusts influence soil enzyme activity (Miralles et al., 2012), nitrification (Castillo-Monroy and Maestre, 2011), and runoff-infiltration rates (Zaady et al., 2013), all of which contribute to soil multifunctionality. Remote sensing provides a valuable and reliable method for mapping biocrusts. Nevertheless, due to the spectral resemblance between predominant dryland surface elements and biocrusts, it is necessary to utilize mapping indices based on hyperspectral data to identify areas dominated by biocrusts at the ecosystem level accurately (Rodríguez-Caballero et al., 2017). This limitation hinders the ability of most satellite imagery products, such as Sentinel, Landsat, or MODIS, to effectively detect biocrusts (Rozenstein and Adamowski, 2017). Because of this, we could not consider biocrusts explicitly in our analyses. However, they have been shown to influence the soil functions we evaluated in drylands significantly (Bowker et al., 2011), and, as such, they could have also influenced our results. Nevertheless, we don´t expect biocrusts to invalidate our results for two main reasons: i) we measured soil functions at 0-7.5 cm depth, and biocrusts affect soil functions largely at the 0-2 cm depth (Pointing and Belnap, 2012), and ii) the positive impacts of perennial vegetation on soil functions such as those studied here extend beyond plant canopies to influence adjacent open areas devoid of perennial vegetation (Maestre et al., 2009).

This study demonstrates the potential of Landsat images and EAM-based models to assess soil multifunctionality over large areas, but several limitations must be acknowledged. Firstly, the temporal resolution of the sensor (one or two images per month) limits the estimations to monthly or yearly intervals, and advanced filters cannot be applied to select images with similar weather conditions within the same month. Secondly, the spectral resolution of the images, with spectral bands of ~30 nm on average in the VIS-NIR region, restricts the quantification of biocrusts, as discussed above, and of critical biophysical variables that evaluate the status of dryland ecosystems, such as the chlorophyll fluorescence or pigment contents of vegetation (Smith et al., 2018; Y. Zhang et al., 2016). Thirdly, the spatial resolution of the images, with pixels of 30 × 30 m, cannot capture the fine-scale spatial heterogeneity that characterizes dryland ecosystems (Smith et al., 2019), as well as that of biocrusts (Maestre and Cortina, 2002). However, new satellite missions will overcome some of these limitations. For instance, Sentinel 2 provides 13 spectral bands and a spatial resolution ranging from 10 m to 60 m, the NASA mission EMIT provides hyperspectral data from 400 nm to 2500 nm with a daily temporal resolution and a spatial resolution of 5 m, and the enhanced spectral resolution of the upcoming Landsat next missions. In addition, the future satellite mission FLEX will provide a single platform of a fluorescence-dedicated imager at an unprecedented spatial resolution of 300m (Meng et al., 2022). The implementation of these new missions will enhance the ability to seamlessly integrate field data, such as those used in this study, with high-resolution indicators of photosynthetic activity and soil properties, such as texture, organic carbon, and moisture. This will improve the accuracy of global models for soil multifunctionality.

Conclusions

The combined use of a unique global field dataset including 14 soil functions and a wide range of RSI calculated from Landsat has enabled us to develop a predictive model for soil multifunctionality in drylands based on three RSI: MCARI2, a soil-atmosphere resistant VI; GEMI, an atmospherically resistant VI; and Ts-Ta, a proxy of water stress conditions. Our findings demonstrate that RSI, such as MCARI2, performs better than NDVI. These findings imply that NDVI is more sensitive to the variability of global dryland covers, a crucial factor in developing comprehensive models for soil multifunctionality in sparsely vegetated regions. To the best of our knowledge, our study is the first to use and demonstrate that thermal-based indicators such as Ts-Ta, which are related to the evapotranspiration rate and water deficit, can improve global models of soil multifunctionality in combination with other RSI. Future research to improve our understanding of dryland dynamics should include EAM methods for accurately upscaling ground-based observations. The soil multifunctionality models developed in this study open the possibility of accurately mapping regional- to global-scale essential soil processes at spatiotemporal resolutions relevant to land managers across drylands worldwide.

Acknowledgements

Field data was obtained with the support of the European Research Council (ERC grant agreement 242658 (BIOCOM). Hernández-Clemente R was supported by the Ramón y Cajal program from the Spanish Ministry of Science and Innovation (RYC2020-029187-I). Maestre FT acknowledges support from Generalitat Valenciana (CIDEGENT/2018/041) and the Spanish Ministry of Science and Innovation (EUR2022-134048).

Authors' Contribution statement

Maestre FT and Hernández-Clemente R conceived the study. Hernández-Clemente R designed the theoretical and computational framework, defined the modelling approaches, and took the lead in writing and preparing the manuscript. Hornero A contributed to the analytical framework, image processing and analytical calculations, Gonzalez-Dugo contributed to the thermal data analysis, Berdugo M and Quero JL contributed to the theoretical framework and discussion of the manuscript, Jiménez JC contributed to the thermal image pre-processing, Maestre FT contributed to the theoretical framework, provided the experimental data and had a major role in writing the manuscript.

References

Austin, P.C., Tu, J.V., 2004. Bootstrap Methods for Developing Predictive Models. Am. Stat. 58, 131–137.

Baret, F., Jacquemoud, S., Hanocq, J.F., 1993. The soil line concept in remote sensing. Remote Sens. Rev. 7, 65–82. https://doi.org/10.1080/02757259309532166

Berdugo, M., Delgado-Baquerizo, M., Soliveres, S., Hernández-Clemente, R., Zhao, Y., Gaitán, J.J., Gross, N., Saiz, H., Maire, V., Lehmann, A., Rillig, M.C., Solé, R.V., Maestre, F.T., 2020. Global ecosystem thresholds driven by aridity. Science 367, 787–790. https://doi.org/10.1126/science.aay5958

Berdugo, M., Kéfi, S., Soliveres, S., Maestre, F.T., 2017. Plant spatial patterns identify alternative ecosystem multifunctionality states in global drylands. Nat. Ecol. Evol. 1, 1–10. https://doi.org/10.1038/s41559-016-0003

Bowker, M., Maestre, F., Mau, R., 2013. Diversity and Patch-Size Distributions of Biological Soil Crusts Regulate Dryland Ecosystem Multifunctionality. Ecosystems 16. https://doi.org/10.1007/s10021-013-9644-5

Bowker, M.A., Mau, R.L., Maestre, F.T., Escolar, C., Castillo-Monroy, A.P., 2011. Functional profiles reveal unique ecological roles of various biological soil crust organisms. Funct. Ecol. 25, 787–795. https://doi.org/10.1111/j.1365-2435.2011.01835.x

Burrell, A.L., Evans, J.P., De Kauwe, M.G., 2020. Anthropogenic climate change has driven over 5 million km2 of drylands towards desertification. Nat. Commun. 11, 3853. https://doi.org/10.1038/s41467-020-17710-7

Castillo-Monroy, A.P., Maestre, F., 2011. Biological soil crusts: Recent advances in our knowledge of their structure and ecological function. Rev. Chil. Hist. Nat. 84, 1–21.

Copertino, 2012. Comparison of algorithms to retrieve Land Surface Temperature from LANDSAT-7 ETM+ IR data in the Basilicata Ionian band. Tethys J. Weather Clim. West. Mediterr. https://doi.org/10.3369/tethys.2012.9.03

Delgado-Baquerizo, M., Maestre, F.T., Gallardo, A., Bowker, M.A., Wallenstein, M.D., Quero, J.L., Ochoa, V., Gozalo, B., García-Gómez, M., Soliveres, S., García-Palacios, P., Berdugo, M., Valencia, E., Escolar, C., Arredondo, T., Barraza-Zepeda, C., Bran, D., Carreira, J.A., Chaieb, M., Conceição, A.A., Derak, M., Eldridge, D.J., Escudero, A., Espinosa, C.I., Gaitán, J., Gatica, M.G., Gómez-González, S., Guzman, E., Gutiérrez, J.R., Florentino, A., Hepper, E., Hernández, R.M., Huber-Sannwald, E., Jankju, M., Liu, J., Mau, R.L., Miriti, M., Monerris, J., Naseri, K., Noumi, Z., Polo, V., Prina, A., Pucheta, E., Ramírez, E., Ramírez-Collantes, D.A., Romão, R., Tighe, M., Torres, D., Torres-Díaz, C., Ungar, E.D., Val, J., Wamiti, W., Wang, D., Zaady, E., 2013. Decoupling of soil nutrient cycles as a function of aridity in global drylands. Nature 502, 672–676. https://doi.org/10.1038/nature12670

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36, 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Elbert, W., Weber, B., Burrows, S., Steinkamp, J., Büdel, B., Andreae, M.O., Pöschl, U., 2012. Contribution of cryptogamic covers to the global cycles of carbon and nitrogen. Nat. Geosci. 5, 459–462. https://doi.org/10.1038/ngeo1486

Fatiha, B., Abdelkader, A., Latifa, H., Mohamed, E., 2013. Spatio Temporal Analysis of Vegetation by Vegetation Indices from Multi-dates Satellite Images: Application to a Semi Arid Area in ALGERIA. Energy Procedia, TerraGreen 13 International Conference 2013 - Advancements in Renewable Energy and Clean Environment 36, 667–675. https://doi.org/10.1016/j.egypro.2013.07.077

Franklin, J., Miller, J.A., 2010. Mapping species distributions: Spatial inference and prediction. Cambridge University Press. https://doi.org/10.1017/CBO9780511810602

Freedman, D. A., 2009. Statistical Models: Theory and Practice, Cambridge University Press. ed. Cambridge University Press, United States of America.

Gaitán, J.J., Bran, D., Oliva, G., Ciari, G., Nakamatsu, V., Salomone, J., Ferrante, D., Buono, G., Massara, V., Humano, G., Celdrán, D., Opazo, W., Maestre, F.T., 2013. Evaluating the performance of multiple remote sensing indices to predict the spatial variability of ecosystem structure and functioning in Patagonian steppes. Ecol. Indic. 34, 181–191. https://doi.org/10.1016/j.ecolind.2013.05.007

Garland, G., Banerjee, S., Edlinger, A., Miranda Oliveira, E., Herzog, C., Wittwer, R., Philippot, L., Maestre, F.T., van der Heijden, M.G.A., 2021. A closer look at the functions behind ecosystem multifunctionality: A review. J. Ecol. 109, 600–613. https://doi.org/10.1111/1365-2745.13511

Gaur, M.K., Squires, V.R., 2017. Climate Variability Impacts on Land Use and Livelihoods in Drylands. Springer.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ., Big Remotely Sensed Data: tools, applications and experiences 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Gosiewska, A., Kozak, A., Biecek, P., 2021. Simpler is better: Lifting interpretability-performance trade-off via automated feature engineering. Decis. Support Syst., Interpretable Data Science For Decision Making 150, 113556. https://doi.org/10.1016/j.dss.2021.113556

Haboudane, D., Miller, J.R., Pattey, E., Zarco-Tejada, P.J., Strachan, I.B., 2004a. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. Remote Sens. Environ. 90, 337–352. https://doi.org/10.1016/j.rse.2003.12.013

Haboudane, D., Miller, J.R., Pattey, E., Zarco-Tejada, P.J., Strachan, I.B., 2004b. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. Remote Sens. Environ. 90, 337–352. https://doi.org/10.1016/j.rse.2003.12.013

Haboudane, D., Miller, J.R., Tremblay, N., Zarco-Tejada, P.J., Dextraze, L., 2002. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. Remote Sens. Environ. 81, 416–426. https://doi.org/10.1016/S0034-4257(02)00018-4

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J.-N., 2020. The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. 146, 1999–2049. https://doi.org/10.1002/qj.3803

Holden, C.E., Woodcock, C.E., 2016. An analysis of Landsat 7 and Landsat 8 underflight data and the implications for time series investigations. Remote Sens. Environ., Landsat 8 Science Results 185, 16–36. https://doi.org/10.1016/j.rse.2016.02.052

Hornero, A., Zarco-Tejada, P.J., Quero, J.L., North, P.R.J., Ruiz-Gómez, F.J., Sánchez-Cuesta, R., Hernandez-Clemente, R., 2021. Modelling hyperspectral- and thermal-based plant traits for the early detection of Phytophthora-induced symptoms in oak decline. Remote Sens. Environ. 263, 112570. https://doi.org/10.1016/j.rse.2021.112570

Huang, J., Ji, M., Xie, Y., Wang, S., He, Y., Ran, J., 2016. Global semi-arid climate change over last 60 years. Clim. Dyn. 46, 1131–1150. https://doi.org/10.1007/s00382-015-2636-8

Huete, A.R., Jackson, R.D., 1987. Suitability of spectral indices for evaluating vegetation characteristics on arid rangelands. Remote Sens. Environ. 23, 213-IN8. https://doi.org/10.1016/0034-4257(87)90038-1

Huete, A.R., Jackson, R.D., Post, D.F., 1985. Spectral response of a plant canopy with different soil backgrounds. Remote Sens. Environ. 17, 37–53. https://doi.org/10.1016/0034-4257(85)90111-7

Jimenez-Muñoz, J.C., Cristobal, J., Sobrino, J.A., Soria, G., Ninyerola, M., Pons, X., Pons, X., 2009. Revision of the Single-Channel Algorithm for Land Surface Temperature Retrieval From Landsat Thermal-Infrared Data. IEEE Trans. Geosci. Remote Sens. 47, 339–349. https://doi.org/10.1109/TGRS.2008.2007125

Kaufman, Y.J., Tanre, D., 1992. Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. IEEE Trans. Geosci. Remote Sens. 30, 261–270. https://doi.org/10.1109/36.134076

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., Hunt, T., 2020. caret: Classification and Regression Training.

Liu, Z., Chen, Z., Yu, G., Zhang, W., Zhang, T., Han, L., 2023. The role of climate, vegetation, and soil factors on carbon fluxes in Chinese drylands. Front. Plant Sci. 14, 1060066. https://doi.org/10.3389/fpls.2023.1060066

Maestre, F.T., Bowker, M.A., Cantón, Y., Castillo-Monroy, A.P., Cortina, J., Escolar, C., Escudero, A., Lázaro, R., Martínez, I., 2011. Ecology and functional roles of biological soil crusts in semi-arid ecosystems of Spain. J. Arid Environ. 75, 1282–1291. https://doi.org/10.1016/j.jaridenv.2010.12.008

Maestre, F.T., Bowker, M.A., Puche, M.D., Belén Hinojosa, M., Martínez, I., García-Palacios, P., Castillo, A.P., Soliveres, S., Luzuriaga, A.L., Sánchez, A.M., Carreira, J.A., Gallardo, A., Escudero, A., 2009. Shrub encroachment can reverse desertification in semi-arid Mediterranean grasslands. Ecol. Lett. 12, 930–941. https://doi.org/10.1111/j.1461-0248.2009.01352.x

Maestre, F.T., Cortina, J., 2002. Spatial patterns of surface soil properties and vegetation in a Mediterranean semi-arid steppe. Plant Soil 241, 279–291. https://doi.org/10.1023/A:1016172308462

Maestre, F.T., Eldridge, D.J., Soliveres, S., Kéfi, S., Delgado-Baquerizo, M., Bowker, M.A., García-Palacios, P., Gaitán, J., Gallardo, A., Lázaro, R., Berdugo, M., 2016. Structure and Functioning of Dryland Ecosystems in a Changing World. Annu. Rev. Ecol. Evol. Syst. 47, 215–237. https://doi.org/10.1146/annurev-ecolsys-121415-032311

Maestre, F.T., Escolar, C., de Guevara, M.L., Quero, J.L., Lázaro, R., Delgado-Baquerizo, M., Ochoa, V., Berdugo, M., Gozalo, B., Gallardo, A., 2013. Changes in biocrust cover drive carbon cycle responses to climate change in drylands. Glob. Change Biol. 19, 3835–3847. https://doi.org/10.1111/gcb.12306

Maestre, F.T., Quero, J.L., Gotelli, N.J., Escudero, A., Ochoa, V., Delgado-Baquerizo, M., García-Gómez, M., Bowker, M.A., Soliveres, S., Escolar, C., García-Palacios, P., Berdugo, M., Valencia, E., Gozalo, B., Gallardo, A., Aguilera, L., Arredondo, T., Blones, J., Boeken, B., Bran, D., Conceição, A.A., Cabrera, O., Chaieb, M., Derak, M., Eldridge, D.J., Espinosa, C.I., Florentino, A., Gaitán, J., Gatica, M.G., Ghiloufi, W., Gómez-González, S., Gutiérrez, J.R., Hernández, R.M., Huang, X., Huber-Sannwald, E., Jankju, M., Miriti, M., Monerris, J., Mau, R.L., Morici, E., Naseri, K., Ospina, A., Polo, V., Prina, A., Pucheta, E., Ramírez-Collantes, D.A., Romão, R., Tighe, M., Torres-Díaz, C., Val, J., Veiga, J.P., Wang, D., Zaady, E., 2012. Plant Species Richness and Ecosystem Multifunctionality in Global Drylands. Science 335, 214–218. https://doi.org/10.1126/science.1215442

Masek, J.G., Vermote, E.F., Saleous, N.E., Wolfe, R., Hall, F.G., Huemmrich, K.F., Gao, F., Kutler, J., Lim, T.-K., 2006. A Landsat surface reflectance dataset for North America, 1990-2000. IEEE Geosci. Remote Sens. Lett. 3, 68–72. https://doi.org/10.1109/LGRS.2005.857030

Meng, X., Cheng, J., Yao, B., Guo, Y., 2022. Validation of the ECOSTRESS Land Surface Temperature Product Using Ground Measurements. IEEE Geosci. Remote Sens. Lett. 19, 1–5. https://doi.org/10.1109/LGRS.2021.3123816

Miralles, I., Ortega, R., Almendros, G., Gil-Sotres, F., Trasar-Cepeda, C., Leirós, M.C., Soriano, M., 2012. Modifications of organic matter and enzymatic activities in response to change in soil use in semi-arid mountain ecosystems (southern Spain). Eur. J. Soil Sci. 63, 272–283. https://doi.org/10.1111/j.1365-2389.2012.01426.x

Morillas, L., García, M., Nieto, H., Villagarcia, L., Sandholt, I., Gonzalez-Dugo, M.P., Zarco-Tejada, P.J., Domingo, F., 2013. Using radiometric surface temperature for surface energy flux estimation in Mediterranean drylands from a two-source perspective. Remote Sens. Environ. 136, 234–246. https://doi.org/10.1016/j.rse.2013.05.010
Ochoa-Hueso, R., Eldridge, D.J., Delgado-Baquerizo, M., Soliveres, S., Bowker, M.A., Gross, N., Le Bagousse-Pinguet, Y., Quero, J.L., García-Gómez, M., Valencia, E., Arredondo, T., Beinticinco, L., Bran, D., Cea, A., Coaguila, D., Dougill, A.J., Espinosa, C.I., Gaitán, J., Guuroh, R.T., Guzman, E., Gutiérrez, J.R., Hernández, R.M., Huber-Sannwald, E., Jeffries, T., Linstädter, A., Mau, R.L., Monerris, J., Prina, A., Pucheta, E., Stavi, I., Thomas, A.D., Zaady, E., Singh, B.K., Maestre, F.T., 2018. Soil fungal abundance and plant functional traits drive fertile island formation in global drylands. J. Ecol. 106, 242–253. https://doi.org/10.1111/1365-2745.12871
Odebiri, O., Mutanga, O., Odindi, J., Naicker, R., 2022. Modelling soil organic carbon stock distribution across different land-uses in South Africa: A remote sensing and deep learning approach. ISPRS J. Photogramm. Remote Sens. 188, 351–362. https://doi.org/10.1016/j.isprsjprs.2022.04.026
Pacheco, J., Casado, S., Porras, S., 2013. Exact methods for variable selection in principal component analysis: Guide functions and pre-selection. Comput. Stat. Data Anal. 57, 95–111. https://doi.org/10.1016/j.csda.2012.06.014
Pinty, B., Verstraete, M.M., 1992. GEMI: a non-linear index to monitor global vegetation from satellites. Vegetatio 101, 15–20. https://doi.org/10.1007/BF00031911
Pointing, S.B., Belnap, J., 2012. Microbial colonization and controls in dryland systems. Nat. Rev. Microbiol. 10, 551–562. https://doi.org/10.1038/nrmicro2831
Prince, S.D., 1991. Satellite remote sensing of primary production: comparison of results for Sahelian grasslands 1981-1988. Int. J. Remote Sens. 12, 1301–1311. https://doi.org/10.1080/01431169108929727
R Core Team, 2013. R: A language and environment for statistical computing.
Ren, H., Feng, G., 2015a. Are soil-adjusted vegetation indices better than soil-unadjusted vegetation indices for above-ground green biomass estimation in arid and semi-arid grasslands? Grass Forage Sci. 70, 611–619. https://doi.org/10.1111/gfs.12152
Ren, H., Feng, G., 2015b. Are soil-adjusted vegetation indices better than soil-unadjusted vegetation indices for above-ground green biomass estimation in arid and semi-arid grasslands? Grass Forage Sci. 70, 611–619. https://doi.org/10.1111/gfs.12152
Rodríguez-Caballero, E., Escribano, P., Olehowski, C., Chamizo, S., Hill, J., Cantón, Y., Weber, B., 2017. Transferability of multi- and hyperspectral optical biocrust indices. ISPRS J. Photogramm. Remote Sens. 126, 94–107. https://doi.org/10.1016/j.isprsjprs.2017.02.007
Rondeaux, G., Steven, M., Baret, F., 1996. Optimization of soil-adjusted vegetation indices. Remote Sens. Environ. 55, 95–107. https://doi.org/10.1016/0034-4257(95)00186-7
Rouse, J.W., Jr., Haas, R.H., Schell, J.A., Deering, D.W., 1974. Monitoring Vegetation Systems in the Great Plains with Erts. NASA Spec. Publ. 351, 309.
Roy, D.P., Wulder, M.A., Loveland, T.R., C.e., W., Allen, R.G., Anderson, M.C., Helder, D., Irons, J.R., Johnson, D.M., Kennedy, R., Scambos, T.A., Schaaf, C.B., Schott, J.R., Sheng, Y., Vermote, E.F., Belward, A.S., Bindschadler, R., Cohen, W.B., Gao, F., Hipple, J.D., Hostert, P., Huntington, J., Justice, C.O., Kilic, A., Kovalskyy, V., Lee, Z.P., Lymburner, L., Masek, J.G., McCorkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne, R.H., Zhu, Z., 2014. Landsat-8: Science and product vision for terrestrial global change research. Remote Sens. Environ. 145, 154–172. https://doi.org/10.1016/j.rse.2014.02.001
Rozenstein, O., Adamowski, J., 2017. A review of progress in identifying and characterizing biocrusts using proximal and remote sensing. Int. J. Appl. Earth Obs. Geoinformation 57, 245–255. https://doi.org/10.1016/j.jag.2017.01.002
Smith, W.K., Biederman, J.A., Scott, R.L., Moore, D.J.P., He, M., Kimball, J.S., Yan, D., Hudson, A., Barnes, M.L., MacBean, N., Fox, A.M., Litvak, M.E., 2018. Chlorophyll Fluorescence Better Captures Seasonal and Interannual Gross Primary Productivity Dynamics Across Dryland Ecosystems of Southwestern North America. Geophys. Res. Lett. 45, 748–757. https://doi.org/10.1002/2017GL075922
Smith, W.K., Dannenberg, M.P., Yan, D., Herrmann, S., Barnes, M.L., Barron-Gafford, G.A., Biederman, J.A., Ferrenberg, S., Fox, A.M., Hudson, A., Knowles, J.F., MacBean, N., Moore, D.J.P., Nagler, P.L., Reed, S.C., Rutherford, W.A., Scott, R.L., Wang, X., Yang, J., 2019. Remote sensing of dryland ecosystem structure and function: Progress, challenges, and opportunities. Remote Sens. Environ. 233, 111401. https://doi.org/10.1016/j.rse.2019.111401

Song, W., Mu, X., Ruan, G., Gao, Z., Li, L., Yan, G., 2017. Estimating fractional vegetation cover and the vegetation index of bare soil and highly dense vegetation with a physically based method. Int. J. Appl. Earth Obs. Geoinformation 58, 168–176. https://doi.org/10.1016/j.jag.2017.01.015

Sonnenschein, R., Kuemmerle, T., Udelhoven, T., Stellmes, M., Hostert, P., 2011. Differences in Landsat-based trend analyses in drylands due to the choice of vegetation estimate. Remote Sens. Environ. 115, 1408–1420. https://doi.org/10.1016/j.rse.2011.01.021

Sun, D., 2015. Detection of dryland degradation using Landsat spectral unmixing remote sensing with syndrome concept in Minqin County, China. Int. J. Appl. Earth Obs. Geoinformation 41, 34–45. https://doi.org/10.1016/j.jag.2015.04.015

Tian, F., Brandt, M., Liu, Y.Y., Verger, A., Tagesson, T., Diouf, A.A., Rasmussen, K., Mbow, C., Wang, Y., Fensholt, R., 2016. Remote sensing of vegetation dynamics in drylands: Evaluating vegetation optical depth (VOD) using AVHRR NDVI and in situ green biomass data over West African Sahel. Remote Sens. Environ. 177, 265–276. https://doi.org/10.1016/j.rse.2016.02.056

Tongway, D.J., Hindley, N., 2004. Landscape Function Analysis Manual : Procedures for Monitoring and Assesssing Landscapes with Special Reference to Minesites and Rangelands. Canberra, CSIRO Sustainable Ecosystems.

Tucker, C.J., Vanpraet, C., Boerwinkel, E., Gaston, A., 1983. Satellite remote sensing of total dry matter production in the Senegalese Sahel. Remote Sens. Environ. 13, 461–474. https://doi.org/10.1016/0034-4257(83)90053-6

UNEP-WCMC, 2007. A spatial analysis approach to the global delineation of dryland areas of relevance to the CBD Programme of Work on Dry and Subhumid Lands. UNEP World Conserv. Monit. Cent.

Vikhar, P.A., 2016. Evolutionary algorithms: A critical review and its future prospects, in: 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC). Presented at the 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC), pp. 261–265. https://doi.org/10.1109/ICGTSPICC.2016.7955308

Wulder, M.A., Roy, D.P., Radeloff, V.C., Loveland, T.R., Anderson, M.C., Johnson, D.M., Healey, S., Zhu, Z., Scambos, T.A., Pahlevan, N., Hansen, M., Gorelick, N., Crawford, C.J., Masek, J.G., Hermosilla, T., White, J.C., Belward, A.S., Schaaf, C., Woodcock, C.E., Huntington, J.L., Lymburner, L., Hostert, P., Gao, F., Lyapustin, A., Pekel, J.-F., Strobl, P., Cook, B.D., 2022. Fifty years of Landsat science and impacts. Remote Sens. Environ. 280, 113195. https://doi.org/10.1016/j.rse.2022.113195

Zaady, E., Arbel, S., Barkai, D., Sarig, S., 2013. Long-term impact of agricultural practices on biological soil crusts and their hydrological processes in a semiarid landscape. J. Arid Environ. 90, 5–11. https://doi.org/10.1016/j.jaridenv.2012.10.021

Zhang, Y., Xiao, X., Jin, C., Dong, J., Zhou, S., Wagle, P., Joanna, J., Guanter, L., Zhang, Yongguang, Zhang, G., Qin, Y., Wang, J., Moore, B., 2016. Consistency between sun-induced chlorophyll fluorescence and gross primary production of vegetation in North America. Remote Sens. Environ. 183, 154–169. https://doi.org/10.1016/j.rse.2016.05.015

Zhang, Z., He, G., Wang, M., Long, T., Wang, G., Zhang, X., 2016. Validation of the generalized single-channel algorithm using Landsat 8 imagery and SURFRAD ground measurements. Remote Sens. Lett. 7, 810–816. https://doi.org/10.1080/2150704X.2016.1190475

Zhao, Y., Wang, X., Novillo, C.J., Arrogante-Funes, P., Vázquez-Jiménez, R., Maestre, F.T., 2018. Albedo estimated from remote sensing correlates with ecosystem multifunctionality in global drylands. J. Arid Environ. 157, 116–123. https://doi.org/10.1016/j.jaridenv.2018.05.010

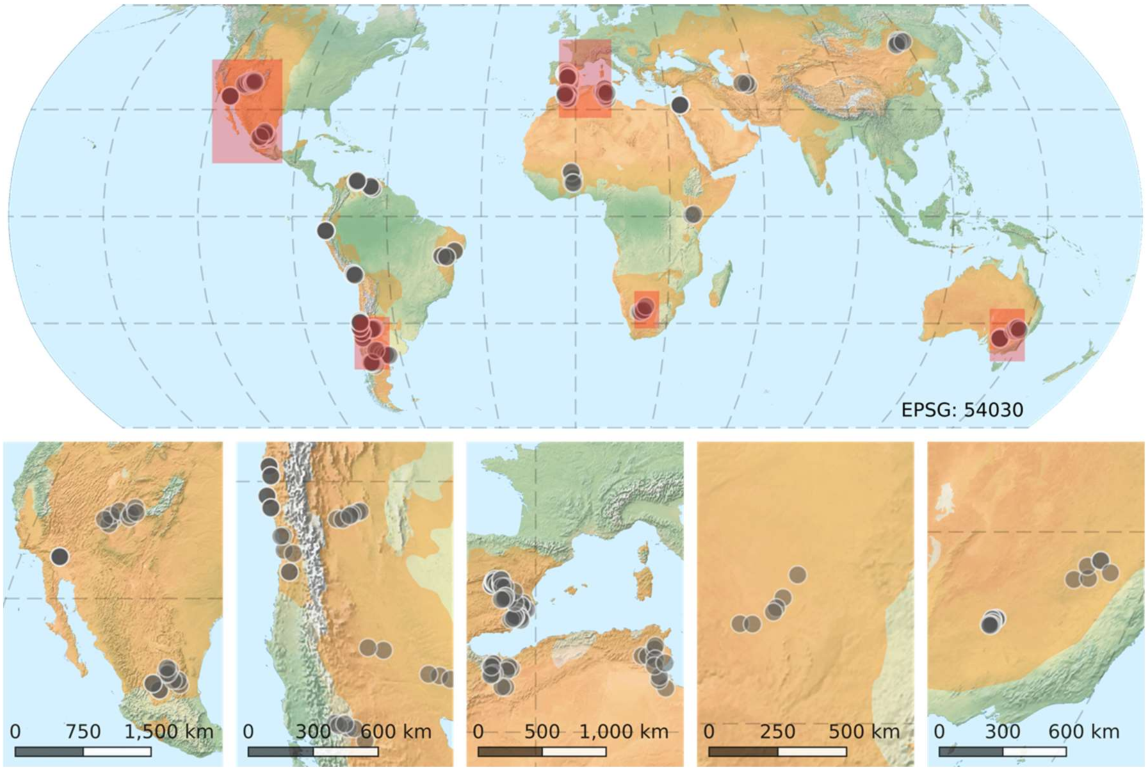


Figure 1. Distribution of the dryland sites used in this study. Dryland land areas are displayed in orange according to FAO/UNEP Land Cover Classification System (UNEP-WCMC, 2007).

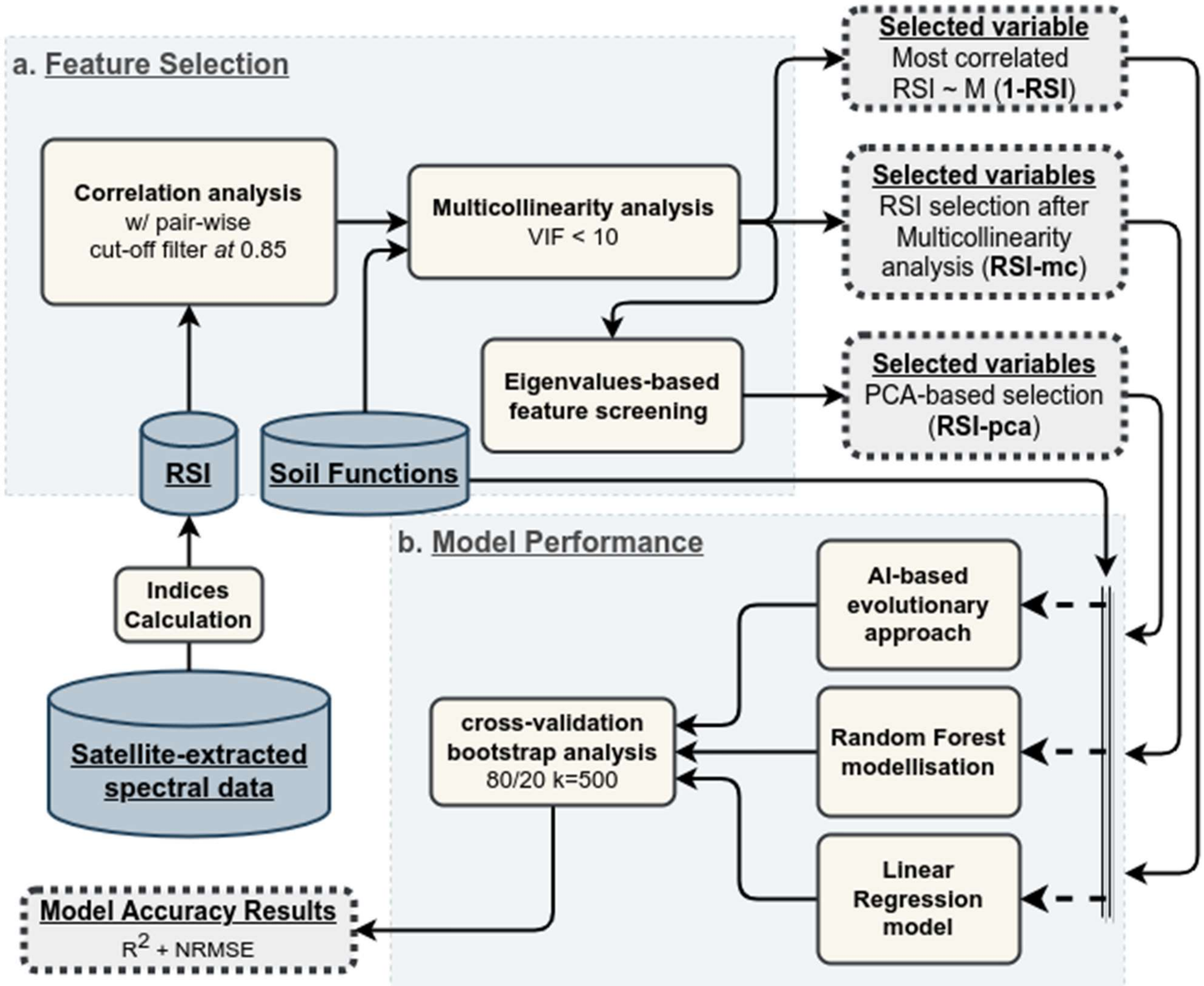


Figure 2. Data analysis workflow – Remote sensing indicators (RSI) selection and model performance evaluation for soil multifunctionality determination.

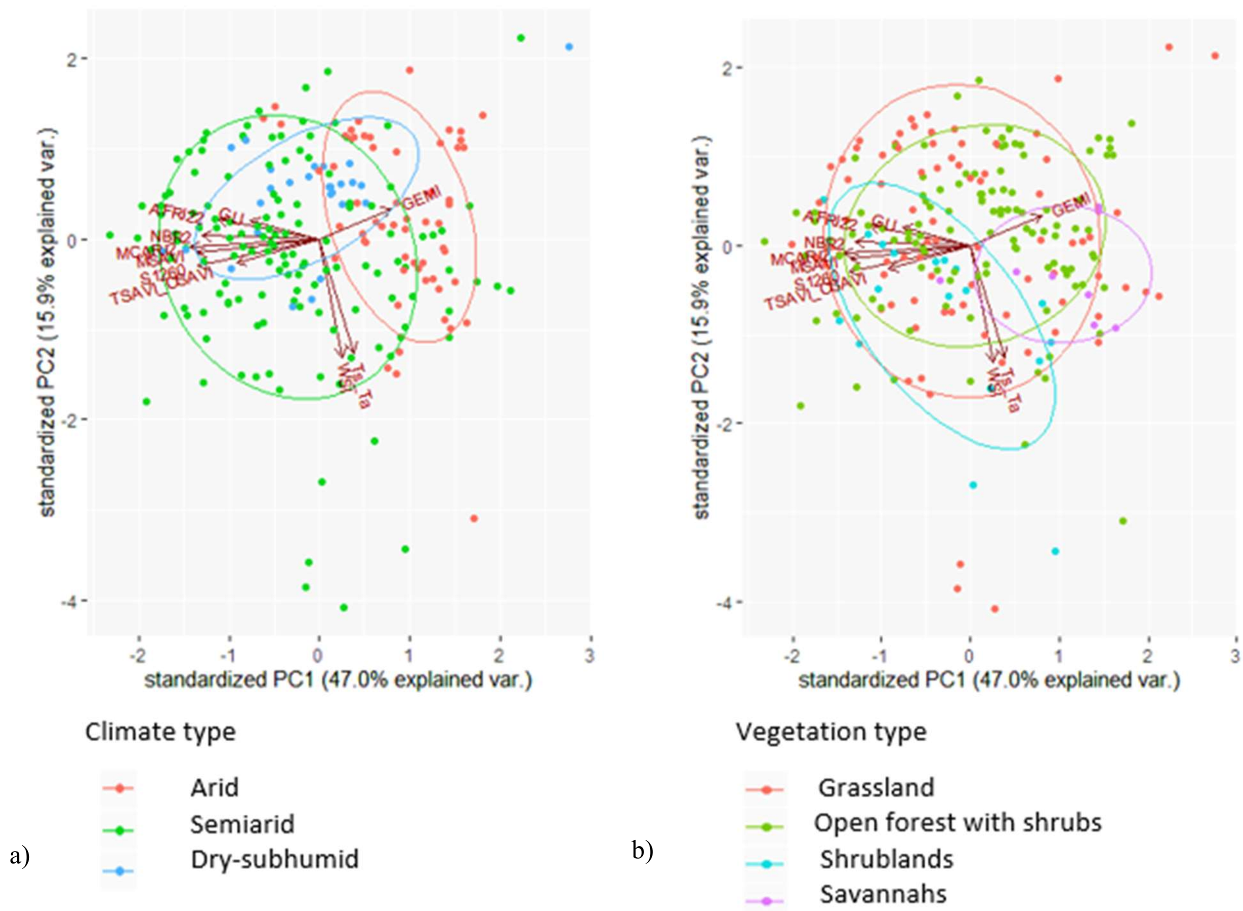


Figure 3. Standardised principal components (PC1 vs PC2) plot of the ten remote sensing indicators (RSI) less correlated among them. Biplot vectors are RSI loadings, whereas the position of the 222 sites is shown within each climate (arid, semi-arid and dry-subhumid, a) and vegetation type (grasslands, shrublands, open forests with shrubs and savannahs, b).

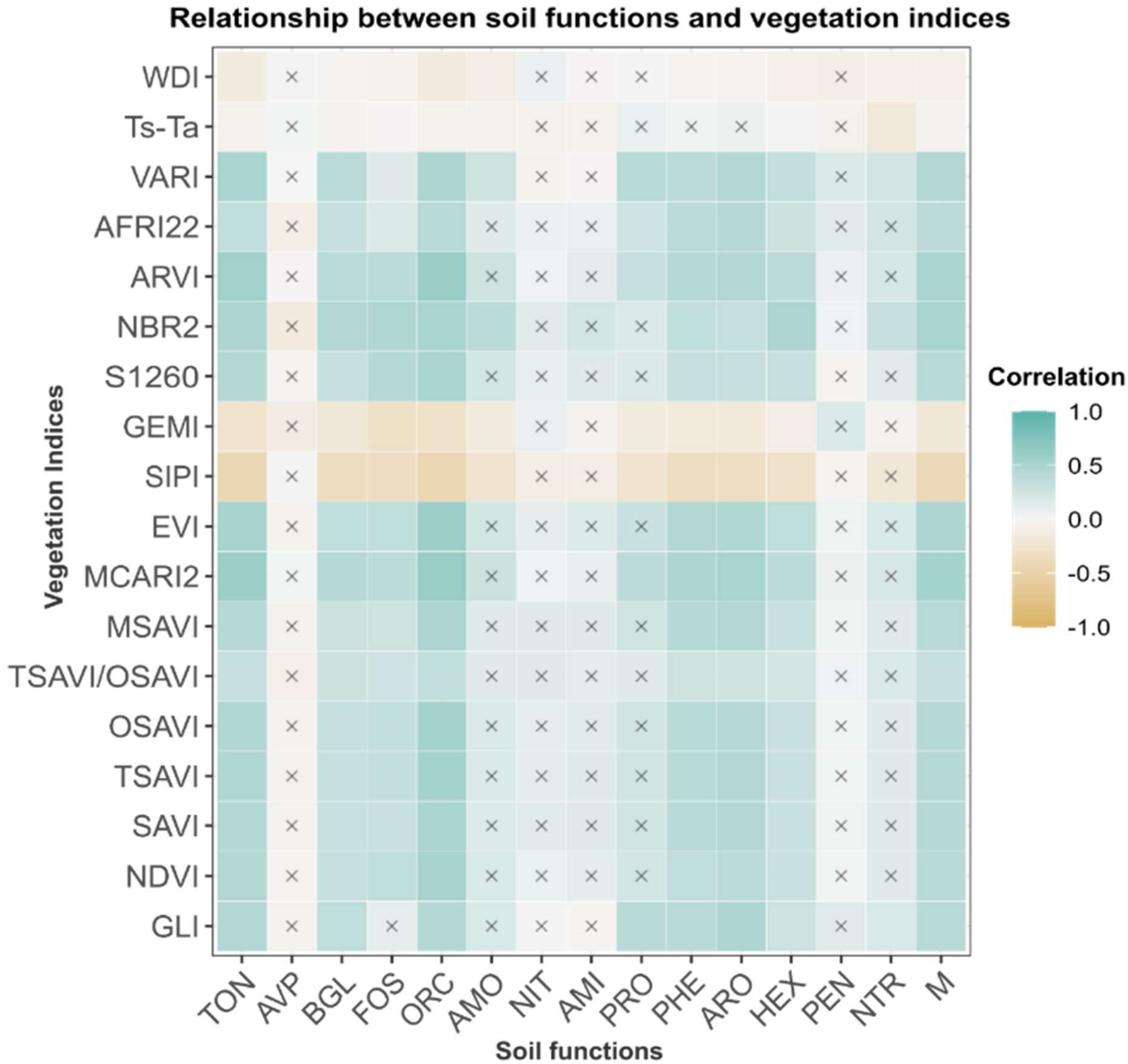


Figure 4. Pearson correlation coefficients between soil multifunctionality (M) and individual soil functions (Table 1) and the different vegetation indices used (Table 2); P > 0.05 value are shown with an "x" symbol (n= 222).

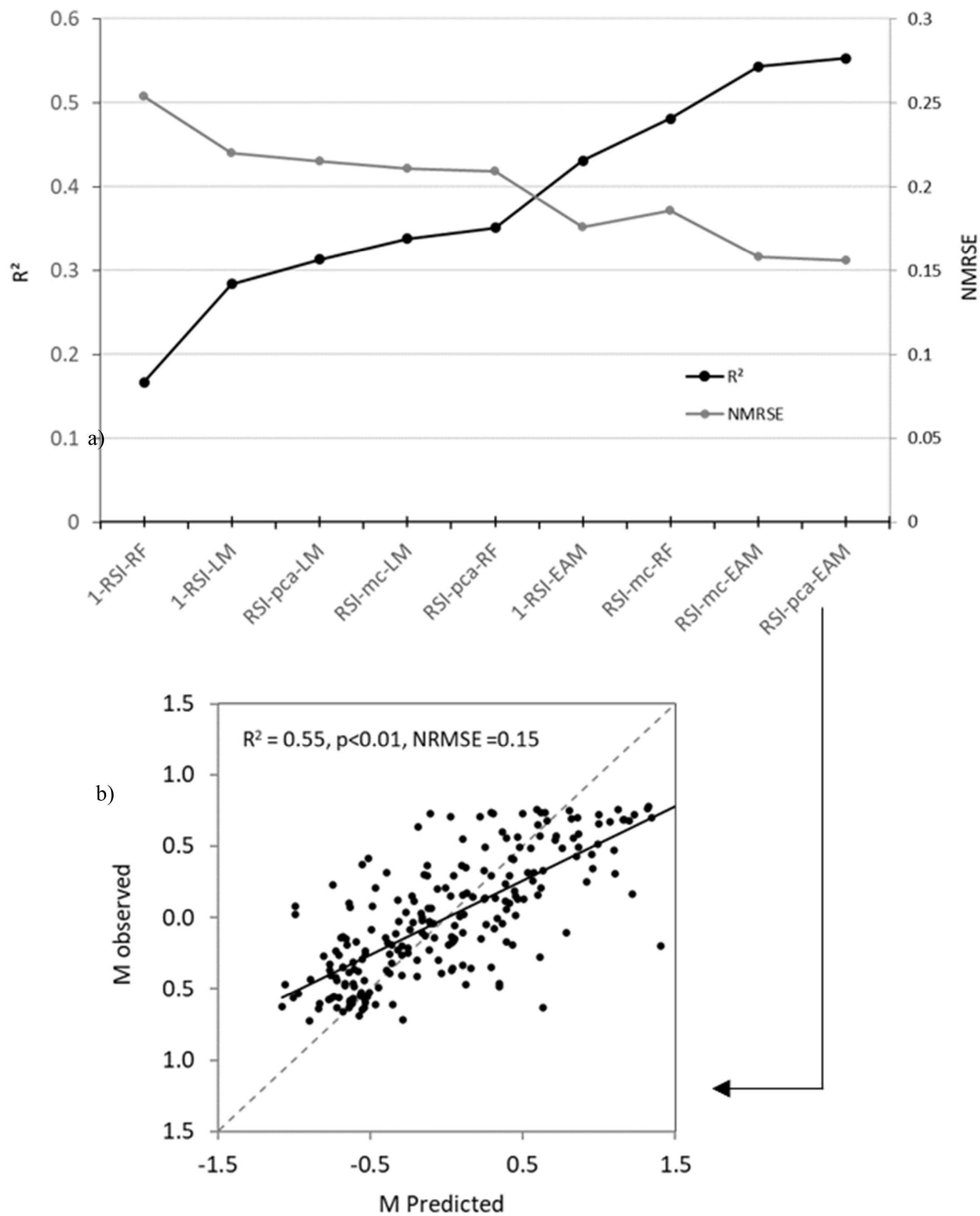
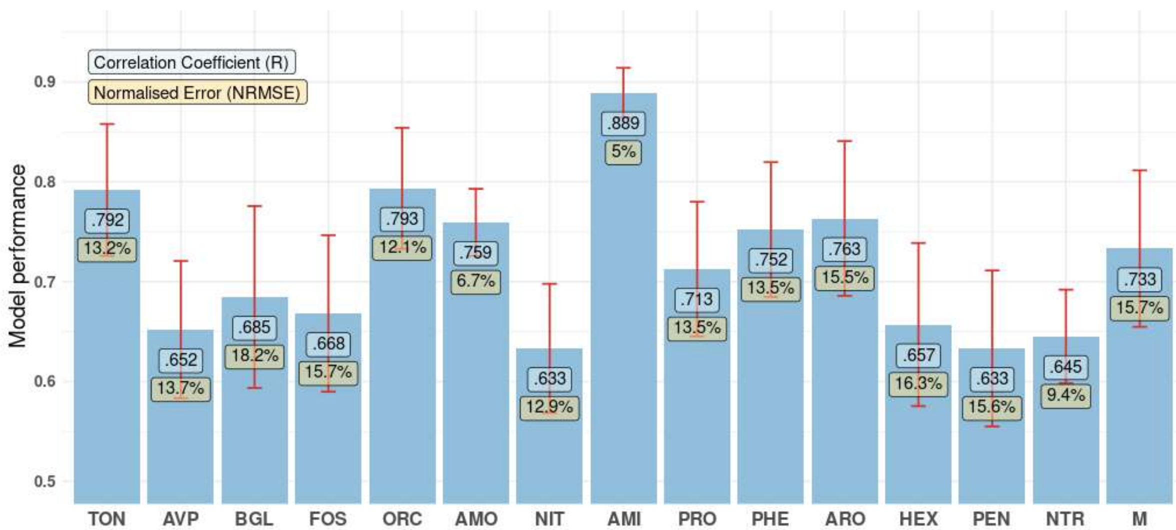


Figure 5. (a) Accuracy results in the prediction of soil multifunctionality according to the coefficient of determination (R^2) and the normalised root mean square error (NMRSE). Results are shown for the linear model (LM), random forest (RF) and evolutionary algorithm model

(EAM) modelling approaches used, and for the three variables reduction sets used: RSI-mc based on MCARI2, NBR2, MSAVI, GLI, S1260, AFRI22, TSAVI_OSAVI, GEMI, Ts-Ta and WDI; RSI-pca based on MCARI2, GEMI and Ts-Ta; and 1-RSI based on MCARI2-. (b) Observed vs predicted soil multifunctionality for the RSI-pca selection with the EAM model using the remote sensing indicators MCARI2, GEMI and Ts-Ta and EAM analysis (n=222). The dashed line represents the 1:1 line. See Table 2 for the acronyms of the indices used.



T

Figure 6. Comparison of the accuracy of EAM-based predictions of soil multifunctionality (M) and soil functions (TON, AVP, BGL, FOS ORC, AMO, NIT, AMI, PRO, PHE, ARO, HEX, PEN and NTR described in Table 1. The results are shown in terms of the correlation coefficient (R) represented with blue columns and the normalised root mean square error (NMRSE) represented with red error bars. The predictions were made using three remote sensing indicators (MCARI2, GEMI and Ts-Ta) selected through principal component analysis (RSI-pca).

Table 1. Soil variables used for the calculation of the soil functions.

Total Nitrogen	TON
Available Potassium	AVP
Activity of b-glucosidase	BGL
Activity of phosphatase	FOS
Organic Carbon	ORC
Ammonium	AMO
Nitrate	NIT
Aminoacids	AMI
Proteins	PRO
Phenols	PHE
Aromatic compounds	ARO
Hexose content	HEX
Pentose content	PEN
Potential N transformation rate	NTR

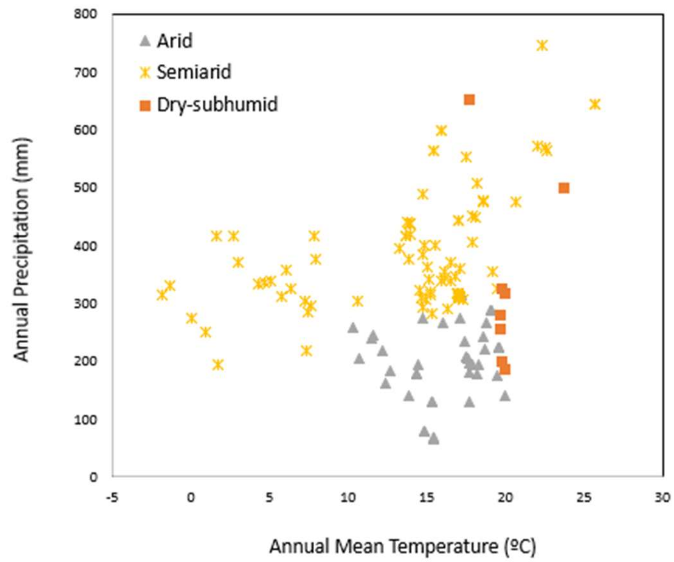
Table 2. Remote sensing indicators and their formulations derived from Landsat data evaluated in this study.

Remote sensing Indicator	Formulation	Reference
GLI: Green Leaf Index	GLI = (2· ρGreen − ρRed − ρBlue)/(2·ρGreen + ρRed + ρBlue)	(Louhaichi et al., 2001)
SIPI: Structure Insensitive Pigment Index	SIPI = (ρNIR - ρBlue) / (ρNIR + ρRed)	(Peñuelas et al., 1995)
NDVI: Normalized Difference Vegetation Index	NDVI = (ρNIR – ρRed)/(ρNIR + ρRed)	(Rouse et al., 1974)
SAVI: Soil-Adjusted Vegetation Index	SAVI = (1+L)*((NIR-Red)/(NIR+Red+L))	(Huete, 1988)
TSAVI: Transformed Soil-Adjusted Vegetation Index	TSAVI = (a*(ρNIR-a*ρRed - b))/(ρRed +a*ρNIR - (a*b)+0.08*(1+a^2))	(Baret and Guyot, 1991)
OSAVI: Optimised Soil-Adjusted Vegetation Index	OSAVI = (ρNIR – ρRed)/(ρNIR + ρRed +0.16)	(Rondeaux et al., 1996)
TSAVI/OSAVI	TSAVI/OSAVI	(Baret and Guyot, 1991)
MSAVI: Modified Soil-Adjusted Vegetation Index	MSAVI = 0.5*(2*ρNIR+1-(((2* ρNIR+1)^ 2)^0.5-8*(ρNIR - ρRed)))	(Qi et al., 1994)
MCARI2: Modified Chlorophyll Absorption Ratio Index 2	MCARI2 = (1.5*(2.5*(ρ800-ρ670)-1.3*(ρ800-R550)))/((2* ρ800+1)^2-(6* ρ800-5*(ρ670)^0.5)-0.5)^0.5	(Haboudane et al., 2002)
EVI: Enhanced Vegetation Index	EVI =2.5*(ρNIR – ρRed)/(ρNIR +6 *ρRed – 7.5* ρBlue +1)	(Huete et al., 2002)
GEMI: Global Environment Monitoring Index	GEMI = n(1-0.25*n)-(Red-0.125)/(1-Red)); n=2(NIR^2-Red^2)+1.5*NIR+0.5*Red)/(NIR+Red+0.5))	(Pinty and Verstraete, 1992)
ARVI: Atmospherically Resistant Vegetation Index	ARVI = (ρNIR -ρRed)-γ(ρRed-ρBlue)/(ρNIR +ρRed)-γ(ρRed-ρBlue)	(Kaufman and Tanre, 1992)
AFRI2100: Aerosol Free Vegetation Index 2100	AFRI2100 = (ρNIR - 0.5*ρ2100)/(ρNIR +0.5*ρ2100)	(Karnieli et al., 2001)
VARI: Visible Atmospherically Resistant Index	VARI = (ρGreen − ρRed)/(ρGreen + ρRed− ρBlue)	(Gitelson et al., 2002)
S1260: Sulphur index 1260	S1260 = (ρ1260-ρ660)/(ρ1260+ρ660)	(Mahajan et al., 2014)
NBR2: Landsat Normalized Burn Ratio 2	NBR2 = (SWIR1 - SWIR2) / (SWIR1 + SWIR2)	(Norton et al., 2009)
Ts-Ta : Surface temperature minus air temperature	Ts-Ta	(Jackson et al., 1981)
WDI: Water Deficit Index	WDI	(Moran et al., 1994)

Supplementary data

Figure S1. Mean annual temperature and mean annual precipitation of the global distribution of drylands classified according to biomes (a) and type of vegetation (b).

a)



b)

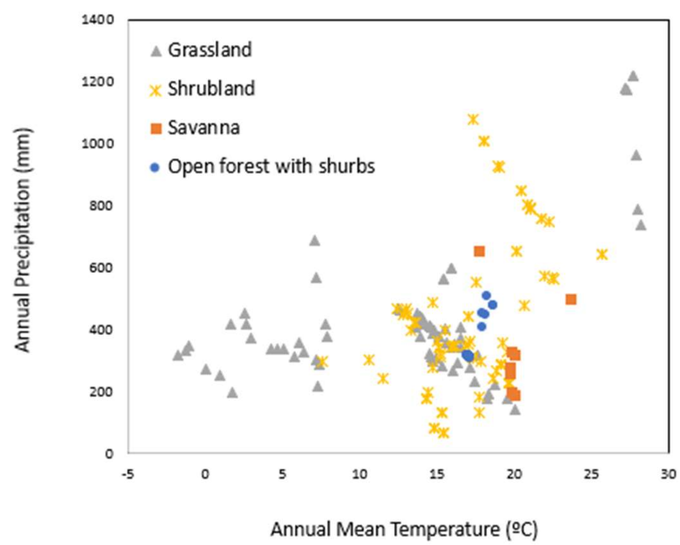


Figure S2. Relationship between the vegetation indices MCARI2 (first line of boxplots) and NDVI (second line of boxplots) with soil multifunctionality (M), the soil functions total nitrogen (TON), and organic carbon (ORC) (n=222) according to the coefficient of determination (R^2) and the correlation coefficient (R) .

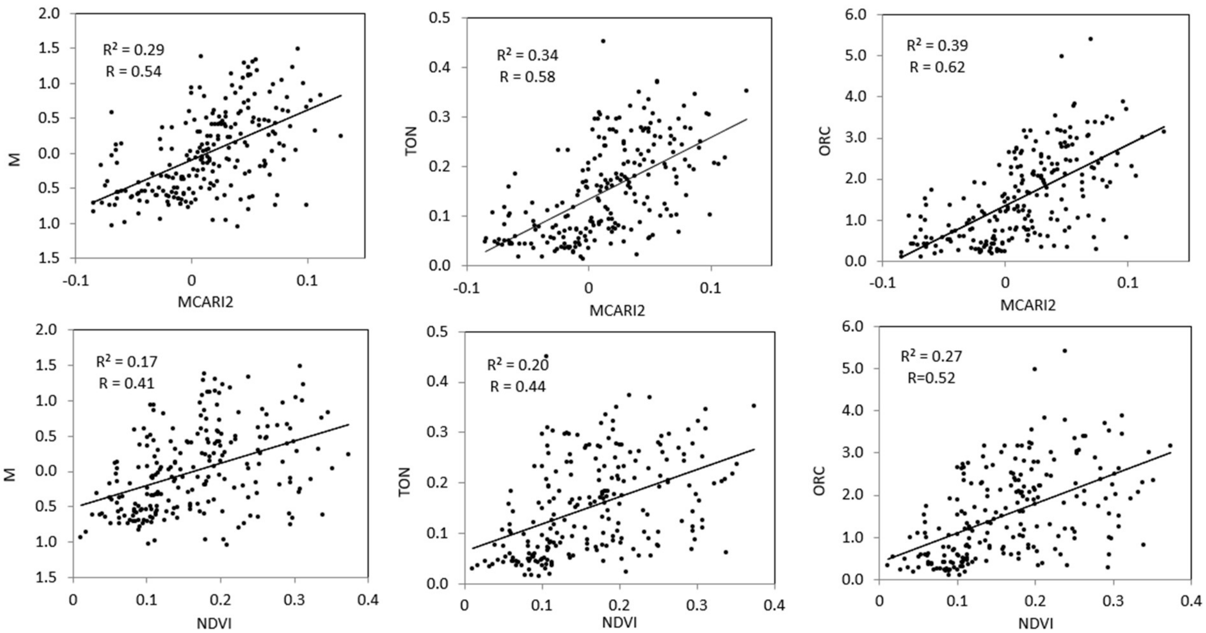

R² = 0.29
R = 0.54
M
MCARI2
R² = 0.34
R = 0.58
TON
MCARI2
R² = 0.39
R = 0.62
ORC
MCARI2
R² = 0.17
R = 0.41
M
NDVI
R² = 0.20
R = 0.44
TON
NDVI
R² = 0.27
R=0.52
ORC
NDVI

Figure S3. Relationships between soil multifunctionality, total cover and the RSI-pc, MCARI2, GEMI and Ts-Ta. Arrow widths are proportional to effect sizes and significance levels. Positive sizes are depicted with green arrows, and negative effects depicted in red.

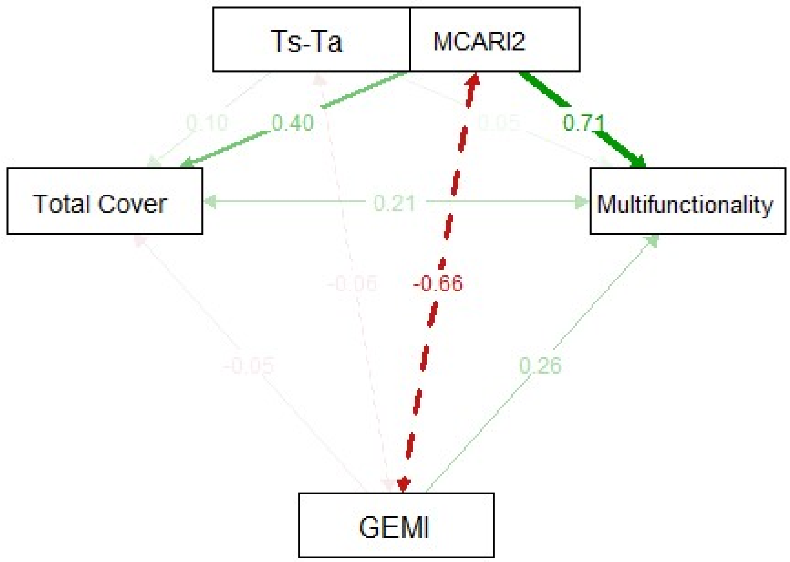


Table S1. Results from the three modelling approaches tested: linear model (LM), random forest (RF) and evolutionary algorithm model (EAM) for the quantification of the soil multifunctionality (M) and soil functions (TON, FOS and ORC) (n=222). Predictors tested include: (RSI-mc) based on MCARI2, NBR2, MSAVI, GLI, S1260, AFRI22, TSAVI_OSAVI, GEMI, Ts-Ta and WDI; (RSI-pc) based on MCARI2, GEMI and Ts-Ta; and (1-RSI) based on MCARI2. The best results for each case are highlighted in bold.

Variable	Method	Predictors	R2	NRMSE	Variable	Method	Predictors	R2	NRMSE
M	LM	RSI-mc	0.338	0.211	TON	LM	RSI-mc	0.388	0.219
M	LM	RSI-pc	0.313	0.215	TON	LM	RSI-pc	0.354	0.222
M	LM	1-RSI	0.284	0.220	TON	LM	1-RSI	0.336	0.225
M	RF	RSI-mc	0.481	0.186	TON	RF	RSI-mc	0.522	0.191
M	RF	RSI-pc	0.351	0.209	TON	RF	RSI-pc	0.352	0.223
M	RF	1-RSI	0.167	0.254	TON	RF	1-RSI	0.210	0.261
M	EAM	RSI-mc	0.543	0.158	TON	EAM	10VI	0.600	0.137
M	**EAM**	**RSI-pc**	**0.553**	**0.156**	**TON**	**EAM**	**RSI-pc**	**0.654**	**0.128**
M	EAM	1-RSI	0.431	0.176	TON	EAM	1-RSI	0.506	0.152
Variable	**Method**	**Predictors**	**R2**	**NRMSE**	**Variable**	**Method**	**Predictors**	**R2**	**NRMSE**
FOS	LM	RSI-mc	0.259	0.221	ORC	LM	RSI-mc	0.397	0.207
FOS	LM	RSI-pc	0.146	0.237	ORC	LM	RSI-pc	0.403	0.206
FOS	LM	1-RSI	0.153	0.236	ORC	LM	1-RSI	0.383	0.209
FOS	RF	RSI-mc	0.330	0.210	ORC	RF	RSI-mc	0.505	0.185
FOS	RF	RSI-pc	0.231	0.225	ORC	RF	RSI-pc	0.408	0.205
FOS	RF	1-RSI	0.080	0.267	ORC	RF	1-RSI	0.262	0.240
FOS	EAM	RSI-mc	0.414	0.156	ORC	EAM	RSI-mc	0.593	0.128
FOS	**EAM**	**RSI-pc**	**0.473**	**0.148**	**ORC**	**EAM**	**RSI-pc**	**0.635**	**0.121**
FOS	EAM	1-RSI	0.379	0.160	ORC	EAM	1-RSI	0.520	0.139

Table S2. Structural equation modelling between soil multifunctionality, total cover and the RSI-pc, MCARI2, GEMI and Ts-Ta. Estimate value, Standard Error, z-values, P(>z), and standardises to the latent factors (Std. lv) and standardised estimates for paths (Std. all).

<table>
<tr><th></th><th>Estimate</th><th>Std.Err</th><th>z-value</th><th>P(>|z|)</th><th>Std. lv</th><th>Std. all</th></tr>
<tr><td>Soil multifunctionality</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>MCARI2</td><td>9.482</td><td>0.990</td><td>9.583</td><td>0.000</td><td>9.482</td><td>0.714</td></tr>
<tr><td>GEMI</td><td>2.269</td><td>0.652</td><td>3.479</td><td>0.001</td><td>2.269</td><td>0.258</td></tr>
<tr><td>Ts_Ta</td><td>0.003</td><td>0.003</td><td>0.907</td><td>0.365</td><td>0.003</td><td>0.051</td></tr>
<tr><td>Total cover</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>MCARI2</td><td>163.173</td><td>33.810</td><td>4.826</td><td>0.000</td><td>163.173</td><td>0.395</td></tr>
<tr><td>GEMI</td><td>-14.421</td><td>22.281</td><td>-0.647</td><td>0.517</td><td>-14.421</td><td>-0.053</td></tr>
<tr><td>Ts_Ta</td><td>0.175</td><td>0.111</td><td>1.570</td><td>0.116</td><td>0.175</td><td>0.097</td></tr>
<tr><td>Covariances:</td><td>Estimate</td><td>Std.Err</td><td>z-value</td><td>P(>|z|)</td><td>Std. lv</td><td>Std. all</td></tr>
<tr><td>Soil multifunctionality</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Total cover</td><td>1.651</td><td>0.548</td><td>3.015</td><td>0.003</td><td>1.651</td><td>0.207</td></tr>
<tr><td>Variances:</td><td>Estimate</td><td>Std.Err</td><td>z-value</td><td>P(>|z|)</td><td>Std. lv</td><td>Std. all</td></tr>
<tr><td>soil multifunctionality</td><td>0.234</td><td>0.022</td><td>10.536</td><td>0.000</td><td>0.234</td><td>0.673</td></tr>
<tr><td>Total cover</td><td>273.094</td><td>25.921</td><td>10.536</td><td>0.000</td><td>273.094</td><td>0.813</td></tr>
</table>